

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5

(i) APPLICANT: Dalla-Favera, Riccardo

(ii) TITLE OF INVENTION: IDENTIFICATION OF GENES ALTERED IN
MULTIPLE MYELOMA

10

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Cooper & Dunham LLP
- (B) STREET: 1185 Avenue of the Americas
- (C) CITY: New York
- (D) STATE: New York
- (E) COUNTRY: U.S.A.
- (F) ZIP: 10036

20

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

25

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE: 28-MAY-1996
- (C) CLASSIFICATION:

30

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: White, John P.
- (B) REGISTRATION NUMBER: 28,678
- (C) REFERENCE/DOCKET NUMBER: 50995

35

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (212) 278-0400
- (B) TELEFAX: (212) 391-0525

40

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 108 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Lys Leu Arg Gln Trp Leu Ile Asp Gln Ile Asp Ser Gly Lys Tyr Pro
1 5 10 15

Gly Leu Val Trp Glu Asn Glu Glu Lys Ser Ile Phe Arg Ile Pro Trp
20 25 30

Lys His Ala Gly Lys Gln Asp Tyr Asn Arg Glu Glu Asp Ala Ala Leu
35 40 45

Phe Lys Ala Trp Ala Leu Phe Lys Gly Lys Phe Arg Glu Gly Ile Asp
50 55 60

Lys Pro Asp Pro Pro Thr Trp Lys Thr Arg Leu Arg Cys Ala Leu Asn
65 70 75 80

30 Lys Ser Asn Asp Phe Glu Glu Leu Val Glu Arg Ser Gln Leu Asp Ile
85 90 95

Ser Asp Pro Tyr Lys Val Tyr Arg Ile Val Pro Glu
35 100 105

(2) INFORMATION FOR SEQ ID NO:2:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

10 Lys Leu Arg Gln Trp Leu Ile Asp Gln Ile Asp Ser Gly Lys Tyr Pro
1 5 10 15

Gly Leu Val Trp Glu Asn Glu Glu Lys Ser Val Phe Arg Ile Pro Trp
20 25 30

15 Lys His Ala Gly Lys Gln Asp Tyr Asn Arg Glu Glu Asp Ala Ala Leu
35 40 45

Phe Lys Ala Trp Ala Leu Phe Lys Gly Lys Phe Arg Glu Gly Ile Asp
20 50 55 60

Lys Pro Asp Pro Pro Thr Trp Lys Thr Arg Leu Arg Cys Ala Leu Asn
65 70 75 80

25 Lys Ser Asn Asp Phe Glu Glu Leu Val Glu Arg Ser Gln Leu Asp Ile
85 90 95

Ser Asp Pro Tyr Lys Val Tyr Arg Ile Val Pro Glu
100 105

30

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
35 (B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Arg Met Arg Pro Trp Leu Glu Met Gln Ile Asn Ser Asn Gln Ile Pro
1 5 10 15

5

Gly Leu Ile Trp Ile Asn Lys Glu Glu Met Ile Phe Gln Ile Pro Trp
20 25 30

Lys His Ala Ala Lys His Gly Trp Asp Ile Asn Lys Asp Ala Cys Leu
10 35 40 45

Phe Arg Ser Trp Ala Ile His Thr Gly Arg Tyr Lys Ala Gly Glu Lys
50 55 60

15 Glu Pro Asp Pro Lys Thr Trp Lys Ala Asn Phe Arg Cys Ala Met Asn
65 70 75 80

Ser Leu Pro Asp Ile Glu Glu Val Lys Asp Gln Lys Arg Asn Lys Gly
85 90 95

20

Ser Ser Ala Val Arg Val Tyr Arg Met Leu Pro Pro
100 105

(2) INFORMATION FOR SEQ ID NO:4:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: peptide

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Arg Met Arg Pro Trp Leu Glu Glu Gln Ile Asn Ser Asn Thr Ile Pro
1 5 10 15

40

Gly Leu Lys Trp Leu Asn Lys Glu Lys Lys Ile Phe Gln Ile Pro Trp

20

25

30

Met His Ala Ala Arg His Gly Trp Asp Val Glu Lys Asp Ala Pro Leu
35 40 45

5

Phe Arg Asn Trp Ala Ile His Thr Gly Lys His Gln Pro Gly Val Asp
50 55 60

Lys Pro Asp Pro Lys Thr Trp Lys Ala Asn Phe Arg Cys Ala Met Asn
10 65 70 75 80

Ser Leu Pro Asp Ile Glu Glu Val Lys Asp Lys Ser Ile Lys Lys Gly
85 90 95

15 Asn Asn Ala Phe Arg Val Tyr Arg Met Leu Pro Leu
100 105

(2) INFORMATION FOR SEQ ID NO:5:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Arg Leu Arg Gln Trp Leu Ile Glu Gln Ile Asp Ser Ser Met Tyr Pro
1 5 10 15

35 Gly Leu Ile Trp Glu Asn Glu Glu Lys Ser Met Phe Arg Ile Pro Trp
20 25 30

40 Lys His Ala Gly Lys Gln Asp Tyr Asn Gln Glu Val Asp Ala Ser Ile
35 40 45

Phe Lys Ala Trp Ala Val Phe Lys Gly Lys Phe Lys Glu Gly Asp Lys

50 55 60

Ala Glu Pro Ala Thr Trp Lys Thr Arg Leu Arg Cys Ala Leu Asn Lys
65 70 75 80

5

Ser Pro Asp Phe Glu Glu Val Thr Asp Arg Ser Gln Leu Asp Ile Ser
85 90 95

Glu Pro Tyr Lys Val Tyr Arg Ile Val Pro Glu
10 100 105

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: peptide

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Lys Leu Arg Asn Trp Val Val Glu Gln Val Glu Ser Gly Gln Phe Pro
1 5 10 15

30 Gly Val Cys Trp Asp Asp Thr Ala Lys Thr Met Phe Arg Ile Pro Trp
20 25 30

Lys His Ala Gly Lys Gln Asp Phe Arg Glu Asp Gln Asp Ala Ala Phe
35 40 45

35 Phe Lys Ala Trp Ala Ile Phe Lys Gly Lys Tyr Lys Glu Gly Asp Thr
50 55 60

40 Gly Gly Pro Ala Val Trp Lys Thr Arg Leu Arg Cys Ala Leu Asn Lys
65 70 75 80

Ser Ser Glu Phe Lys Glu Val Pro Glu Arg Gly Arg Met Asp Val Ala

85

90

95

Glu Pro Tyr Lys Val Tyr Gln Leu Leu Pro Pro
100 105

5

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

20

Arg Ile Leu Pro Trp Leu Val Ser Gln Leu Asp Leu Gly Gln Leu Glu
1 5 10 15

25

Gly Val Ala Trp Val Asn Lys Ser Arg Thr Arg Phe Arg Ile Pro Trp
20 25 30

30

Lys His Gly Leu Arg Gln Asp Ala Gln Gln Glu Asp Phe Gly Ile Phe
35 40 45

35

Gln Ala Trp Ala Glu Ala Thr Gly Ala Tyr Val Pro Gly Arg Asp Lys
50 55 60

Pro Asp Leu Pro Thr Trp Lys Arg Asn Phe Arg Ser Ser Ala Leu Asn
65 70 75 80

40

Arg Lys Glu Gly Leu Arg Leu Ala Glu Asp Arg Ser Lys Asp Pro His
85 90 95

Asp Pro His Lys Ile Tyr Glu Phe Val Asn Ser
100 105

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: peptide

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Lys Arg Leu Cys Gln Ser Thr Ile Tyr Trp Asp Gly Pro Leu Ala Leu
1 5 10 15

15

Cys Asn Asp Arg Pro Asn Lys Leu Glu Arg Asp Gln Thr Cys Lys Leu
20 25 30

20

Phe Asp Thr Gln Gln Phe Leu Ser Glu Leu Gln Ala Phe Ala His His
35 40 45

25

Gly Arg Ser Leu Pro Arg Phe Gln Val Thr Leu Cys Phe Gly Glu Glu
50 55 60

30

Phe Pro Asp Pro Gln Arg Gln Arg Lys Leu Ile Thr Ala His Val Glu
65 70 75 80

Pro Leu Leu Ala Arg Gln Leu Tyr Tyr Phe Ala Gln Gln Asn Ser
85 90 95

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: peptide

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Lys Arg Leu Cys Gln Ser Arg Ile Tyr Trp Asp Gly Pro Leu Ala Leu
5 1 5 10 15

Cys Ser Asp Arg Pro Asn Lys Leu Glu Arg Asp Gln Thr Cys Lys Leu
20 25 30

10 Phe Asp Thr Gln Gln Phe Leu Ser Glu Leu Gln Val Phe Ala His His
35 40 45

Gly Arg Pro Ala Pro Arg Phe Gln Val Thr Leu Cys Phe Gly Glu Glu
50 55 60

15 Phe Pro Asp Pro Gln Arg Gln Arg Lys Leu Ile Thr Ala His Val Glu
65 70 75 80

20 Pro Leu Leu Ala Arg Gln Leu Tyr Tyr Phe Ala Gln Gln Asn Thr
85 90 95

(2) INFORMATION FOR SEQ ID NO:10:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- 30 (ii) MOLECULE TYPE: peptide

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Lys Arg Leu Cys Gln Gly Arg Val Phe Cys Ser Gly Asn Ala Val Val
1 5 10 15

40 Cys Lys Gly Arg Pro Asn Lys Leu Glu Arg Asp Glu Val Val Gln Val
20 25 30

Phe Asp Thr Ser Gln Phe Phe Arg Glu Leu Gln Gln Phe Tyr Asn Ser
35 40 45

Gln Gly Arg Leu Pro Asp Gly Arg Val Val Leu Cys Phe Gly Glu Glu
5 50 55 60

Phe Pro Asp Met Ala Pro Leu Arg Ser Lys Leu Ile Leu Val Gln Ile
65 70 75 80

10 Glu Gln Leu Tyr Val Arg Gln Leu Ala Glu Glu Ala Gly Lys Ser Cys
85 90 95

(2) INFORMATION FOR SEQ ID NO:11:

15

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

30 Gln Arg Leu Cys Pro Ile Pro Ile Ser Trp Asn Ala Pro Gln Ala Pro
1 5 10 15

Pro Gly Pro Gly Pro His Leu Leu Pro Ser Asn Glu Cys Val Glu Leu
20 25 30

35 Phe Arg Thr Ala Tyr Phe Cys Arg Asp Leu Val Arg Tyr Phe Gln Gly
35 40 45

Leu Gly Pro Pro Pro Lys Phe Gln Val Thr Leu Asn Phe Trp Glu Glu
50 55 60

40 Ser His Gly Ser Ser His Thr Pro Gln Asn Leu Ile Thr Val Lys Met
65 70 75 80

Glu Gln Ala Phe Ala Arg Tyr Leu Leu Glu Gln Thr Pro Glu Gln Gln
85 90 95

5 (2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 100 amino acids
(B) TYPE: amino acid
10 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

20 Gln Arg Leu Gly His Cys His Thr Tyr Trp Ala Val Ser Glu Glu Leu
1 5 10 15

Leu Pro Asn Ser Gly His Gly Pro Asp Gly Glu Val Pro Lys Asp Lys
20 25 30

25 Glu Gly Gly Val Phe Asp Leu Gly Pro Phe Ile Val Asp Leu Ile Thr
35 40 45

30 Phe Thr Glu Gly Ser Gly Arg Ser Pro Arg Tyr Ala Trp Leu Phe Cys
50 55 60

Val Gly Glu Ser Trp Pro Gln Asp Gln Pro Trp Thr Lys Arg Leu Val
65 70 75 80

35 Met Val Lys Val Val Pro Thr Cys Leu Arg Ala Leu Val Glu Met Ala
85 90 95

Arg Val Gly Gly

100

40

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5176 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: other nucleic acid

10

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 217..1569

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCCTGACCAA CATGGTAAAA CCCCATCTCT GCTAAACTA CAAAAAATTA GCTGGATGTG

60

20

GTGGCAGGGA ACCTGTCATC CCAGCTAGTT GGGAGACTGA GGCAGGAGAA TCGCTCGATC

120

25

TTGGGACCCA CCGCTGCCCT CAGCTCCGAG TCCAGGGCGA GTGCAGAGCA CAGCGGGCGG

180

AGGACCCCCGG GCGCGGGCGC GGACGGCACCG CGGGGC ATG AAC CTG GAG GGC GGC

234

Met Asn Leu Glu Gly Gly

1

5

30

GGC CGA GGC GGA GAG TTC GGC ATG AGC GCG GTG AGC TGC GGC AAC GGG

282

Gly Arg Gly Gly Glu Phe Gly Met Ser Ala Val Ser Cys Gly Asn Gly

10

15

20

35

AAG CTC CGC CAG TGG CTG ATC GAC CAG ATC GAC AGC GGC AAG TAC CCC

330

Lys Leu Arg Gln Trp Leu Ile Asp Gln Ile Asp Ser Gly Lys Tyr Pro

25

30

35

40

GGG CTG GTG TGG GAG AAC GAG GAG AAG AGC ATC TTC CGC ATC CCC TGG

378

Gly Leu Val Trp Glu Asn Glu Glu Lys Ser Ile Phe Arg Ile Pro Trp
40 45 50

AAG CAC GCG GGC AAG CAG GAC TAC AAC CGC GAG GAG GAC GCC GCG CTC
5 426
Lys His Ala Gly Lys Gln Asp Tyr Asn Arg Glu Glu Asp Ala Ala Leu
55 60 65 70

TTC AAG GCT TGG GCA CTG TTT AAA GGA AAG TTC CGA GAA GGC ATC GAC
10 474
Phe Lys Ala Trp Ala Leu Phe Lys Gly Lys Phe Arg Glu Gly Ile Asp
75 80 85

AAG CCG GAC CCT CCC ACC TGG AAG ACG CGC CTG CGG TGC GCT TTG AAC
15 522
Lys Pro Asp Pro Pro Thr Trp Lys Thr Arg Leu Arg Cys Ala Leu Asn
90 95 100

AAG AGC AAT GAC TTT GAG GAA CTG GTT GAG CGG AGC CAG CTG GAC ATC
20 570
Lys Ser Asn Asp Phe Glu Glu Leu Val Glu Arg Ser Gln Leu Asp Ile
105 110 115

TCA GAC CCG TAC AAA GTG TAC AGG ATT GTT CCT GAG GGA GCC AAA AAA
25 618
Ser Asp Pro Tyr Lys Val Tyr Arg Ile Val Pro Glu Gly Ala Lys Lys
120 125 130

GGA GCC AAG CAG CTC ACC CTG GAG GAC CCG CAG ATG TCC ATG AGC CAC
30 666
Gly Ala Lys Gln Leu Thr Leu Glu Asp Pro Gln Met Ser Met Ser His
135 140 145 150

CCC TAC ACC ATG ACA ACG CCT TAC CCT TCG CTC CCA GCC CAG CAG GTT
35 714
Pro Tyr Thr Met Thr Pro Tyr Pro Ser Leu Pro Ala Gln Gln Val
155 160 165

CAC AAC TAC ATG ATG CCA CCC CTC GAC CGA AGC TGG AGG GAC TAC GTC
40 762
His Asn Tyr Met Met Pro Pro Leu Asp Arg Ser Trp Arg Asp Tyr Val
170 175 180

CCG GAT CAG CCA CAC CCG GAA ATC CCG TAC CAA TGT CCC ATG ACG TTT
810
Pro Asp Gln Pro His Pro Glu Ile Pro Tyr Gln Cys Pro Met Thr Phe
185 190 195

5
GGA CCC CGC GGC CAC CAC TGG CAA GGC CCA GCT TGT GAA AAT GGT TGC
858
Gly Pro Arg Gly His His Trp Gln Gly Pro Ala Cys Glu Asn Gly Cys
200 205 210

10
CAG GTG ACA GGA ACC TTT TAT GCT TGT GCC CCA CCT GAG TCC CAG GCT
906
Gln Val Thr Gly Thr Phe Tyr Ala Cys Ala Pro Pro Glu Ser Gln Ala
215 220 225 230

15
CCC GGA GTC CCC ACA GAG CCA AGC ATA AGG TCT GCC GAA GCC TTG GCG
954
Pro Gly Val Pro Thr Glu Pro Ser Ile Arg Ser Ala Glu Ala Leu Ala
235 240 245

20
TTC TCA GAC TGC CGG CTG CAC ATC TGC CTG TAC TAC CGG GAA ATC CTC
1002
Phe Ser Asp Cys Arg Leu His Ile Cys Leu Tyr Tyr Arg Glu Ile Leu
250 255 260

25
GTG AAG GAG CTG ACC ACG TCC AGC CCC GAG GGC TGC CGG ATC TCC CAT
1050
Val Lys Glu Leu Thr Thr Ser Ser Pro Glu Gly Cys Arg Ile Ser His
265 270 275

30
GGA CAT ACG TAT GAC GCC AGC AAC CTG GAC CAG GTC CTG TTC CCC TAC
1098
Gly His Thr Tyr Asp Ala Ser Asn Leu Asp Gln Val Leu Phe Pro Tyr
280 285 290

35
CCA GAG GAC AAT GGC CAC AGG AAA AAC ATT GAG AAC CTG CTG AGC CAC
1146
Pro Glu Asp Asn Gly His Arg Lys Asn Ile Glu Asn Leu Leu Ser His
295 300 305 310

40
CTG GAG AGG GGC GTG GTC CTC TGG ATG GCC CCC GAC GGG CTC TAT GCG
1194

Leu Glu Arg Gly Val Val Leu Trp Met Ala Pro Asp Gly Leu Tyr Ala
315 320 325

AAA AGA CTG TGC CAG AGC ACG ATC TAC TGG GAC GGG CCC CTG GCG CTG
5 1242

Lys Arg Leu Cys Gln Ser Thr Ile Tyr Trp Asp Gly Pro Leu Ala Leu
330 335 340

TGC AAC GAC CGG CCC AAC AAA CTG GAG AGA GAC CAG ACC TGC AAG CTC
10 1290

Cys Asn Asp Arg Pro Asn Lys Leu Glu Arg Asp Gln Thr Cys Lys Leu
345 350 355

TTT GAC ACA CAG CAG TTC TTG TCA GAG CTG CAA GCG TTT GCT CAC CAC
15 1338

Phe Asp Thr Gln Gln Phe Leu Ser Glu Leu Gln Ala Phe Ala His His
360 365 370

GGC CGC TCC CTG CCA AGA TTC CAG GTG ACT CTA TGC TTT GGA GAG GAG
20 1386

Gly Arg Ser Leu Pro Arg Phe Gln Val Thr Leu Cys Phe Gly Glu Glu
375 380 385 390

TTT CCA GAC CCT CAG AGG CAA AGA AAG CTC ATC ACA GCT CAC GTA GAA
25 1434

Phe Pro Asp Pro Gln Arg Gln Arg Lys Leu Ile Thr Ala His Val Glu
395 400 405

CCT CTG CTA GCC AGA CAA CTA TAT TAT TTT GCT CAA CAA AAC AGT GGA
30 1482

Pro Leu Leu Ala Arg Gln Leu Tyr Tyr Phe Ala Gln Gln Asn Ser Gly
410 415 420

CAT TTC CTG AGG GGC TAC GAT TTA CCA GAA CAC ATC AGC AAT CCA GAA
35 1530

His Phe Leu Arg Gly Tyr Asp Leu Pro Glu His Ile Ser Asn Pro Glu
425 430 435

GAT TAC CAC AGA TCT ATC CGC CAT TCC TCT ATT CAA GAA TGAAAAATGT
40 1579

Asp Tyr His Arg Ser Ile Arg His Ser Ser Ile Gln Glu
440 445 450

CAAGATGAGT GGTTTCTTT TTCCCTTTTT TTTTTTTTT TTTTGATACG GAGATACGGG
1639

5 GTCTTGCTCT GTCTCCCAGG CTGGAGTGCA GTGACACAAT CTCAGCTCAC TGTGACCTCC
1699

GCCTCCTGGG TTCAAGAGAC TCTCCTGCCT CAGCCTCCCT GGTAGCTGGG ATTACAGGTG
1759

10 TGAGCCACTG CACCCACCCA AGACAAGTGA TTTTCATTGT AAATATTGA CTTTAGTGAA
1819

AGCGTCCAAT TGACTGCCCT CTTACTGTTT TGAGGAACTC AGAAGTGGAG ATTCAGTTC
1879

15 AGCGGTTGAG GAGAATTGCG GCGAGACAAG CATGGAAAAT CAGTGACATC TGATTGGCAG
1939

ATGAGCTTAT TTCAAAAGGA AGGGTGGCTT TGCACTTCT TGTGTTCTGT AGACTGCCAT
20 1999

CATTGATGAT CACTGTAAA ATTGACCAAG TGATGTGTTT ACATTTACTG AAATGCGCTC
2059

25 TTTAATTGT TGTAGATTAG GTCTTGCTGG AAGACAGAGA AAACTTGCCT TTCAGTATTG
2119

ACACTGACTA GAGTGATGAC TGCTTGTAGG TATGTCTGTG CCATTTCTCA GGGAAAGTAAG
2179

30 ATGTAAATTG AAGAACGCTC ACACGTAAAA GAAATGTATT AATGTATGTA GGAGCTGCAG
2239

TTCTTGTGGA AGACACTTGC TGAGTGAAGG AAATGAATCT TTGACTGAAG CCGTGCCTGT
35 2299

AGCCTTGGGG AGGCCCATCC CCCACCTGCC AGCGGTTCC TGGTGTGGGT CCCTCTGCC
2359

40 CACCCTCCTT CCCATTGGCT TTCTCTCCTT GGCTTCCCT GGAAGCCAGT TAGTAAACTT
2419

CCTATTTCT TGAGTCAAA AACATGAGCG CTACTCTGG ATGGGACATT TTTGTCTGTC
2479

5 CTACAATCTA GTAATGTCTA AGTAATGGTT AAGTTTCTT GTTTCTGCAT CTTTTTGACC
2539

CTCATTCTT AGAGATGCTA AAATTCTTCG CATAAAGAAG AAGAAATTAA GGAACATAAA
2599

10 TCTTAATACT TGAACGTGG CCCTCTGTC CAAGTACTTA ACTATCTGTT CCCTTCCTCT
2659

GTCGCCACGCT CCTCTGTTG TTTGGCTGTC CAGCGATCAG CCATGGCGAC ACTAAAGGAG
2719

15 GAGGAGCCGG GGACTCCCAG GCTGGAGAGC ACTGCCAGGA CCCACCCTG GAAGCAGGAT
2779

20 GGAGCTGACT ACGGAACCTGC ACACTCAGTG GGCTGTTCT GCTTATTCA TCTGTTCTAT
2839

GCTTCCTCGT GCCAATTATA GTTTGACAGG GCCTTAAAT TACTTGGCTT TTTCCAAATG
2899

25 CTTCTATTAA TAGAAATCCC AAAGACCTCC ACTTGCTTAA GTATACCTAT CACTTACATT
2959

30 TTTGTGGTT TGAGAAAGTA CAGCAGTAGA CTGGGGCGTC ACCTCCAGGC CGTTTCTCAT
3019

ACTACAGGAT ATTTACTATT ACTCCCAGGA TTCAGCAGAA GATTGCGTTA GCTCTCAAAT
3079

35 GTGTGTTCT GCTTTCTAA TGGATATTTT AAATTCATTC AACAAAGCACC TAGTAAGTGC
3139

CTGCTGTATC CCTACATTAC ACAGTTCAGC CTTTATCAAG CTTAGTGAGC AGTGAGCACT
3199

40 GAAACATTAT TTTTTAATGT TTAAAAAGTT TCTAATATTA AAGTCAGAAT ATTAATACAA
3259

TTAATATTAA TATTAAC TAC AGAAAAGACA AACAGTAGAG AACAGCAAA AAATAAAAAG
3319

5 GATCTCCTTT TTTCCCAGCC CAAATTCTCC TCTCTAAAAG TGTCCACAAG AAGGGGTGTT
3379

TATTCTTCCA ACACATTCA CTTTCTGTA AATATACATA AACTTAAAAA GAAAACCTCA
3439

10 TGGAGTCATC TTGCACACAC TTTTCATGCA GTGCTCTTG TAGCTAAACA GTGAAGATT
3499

ACCTCGTTCT GCTCAGAGGC CTTGCTGTGG AGCTCCACTG CCATGTACCC AGTAGGGTTT
3559

15 GACATTTCAT TAGCCATGCA ACATGGATAT GTATTGGCA GCAGACTGTG TTTCGTGAAC
3619

20 TGCAGTGATG TATACATCTT ATAGATGCAA AGTATTTGG GGTATATTAT CCTAAGGGAA
3679

GATAAAAGATG ATATTAAGAA CTGCTGTTTC ACGGGGCCCT TACCTGTGAC CCTCTTGCT
3739

25 GAAGAATATT TAACCCCACA CAGCACTTCA AAGAAGCTGT CTTGGAAGTC TGTCTCAGGA
3799

30 GCACCCTGTC TTCTTAATTC TCCAAGCGGA TGCTCCATT CAATTGCTTT GTGACTTCTT
3859

CTTCTTGTT TTTTAAATA TTATGCTGCT TTAACAGTGG AGCTGAATTT TCTGGAAAAT
3919

35 GCTTCTTGGC TGGGGCCACT ACCTCCTTTC CTATCTTAC ATCTATGTGT ATGTTGACTT
3979

TTTAAAATTC TGAGTGATCC AGGGTATGAC CTAGGGAATG AACTAGCTAT GGAAATAACT
4039

40 CAGGGTTAGG AATCCTAGCA CTTGTCTCAG GACTCTGAAA AGGAACGGCT TCCTCATTCC
4099

TTGTCTTGAT AAAGTGGAAAT TGGCAAACCA GAATTTAGTT TGTACTCAGT GGACAGTGCT
4159

GTTGAAGATT TGAGGACTTG TTAAAGAGCA CTGGGTCATA TGGAAAAAAT GTATGTGTCT
5 4219

CCCCCAGGTGC ATTTTCTTGG TTTATGTCTT GTTCTTGAGA TTTTGTATAT TTAGGAAAAC
4279

10 CTCAAGCAGT AATTAATATC TCCTGGAACA CTATAGAGAA CCAAGTGACC GACTCATTAA
4339

CAACTGAAAC CTAGGAAGCC CCTGAGTCCT GAGCGAAAAC AGGAGAGTTA GTGCCCTAC
4399

15 AGAAAAACCA GCTAGACTAT TGGGTATGAA CTAAAAGAG ACTGTGCCAT GGTGAGAAAA
4459

ATGTAAAATC CTACAGTGGA ATGAGCAGCC CTTACAGTGT TGTTACCACC AAGGGCAGGT
20 4519

AGGTATTAGT GTTTGAAAAA GCTGGTCTTT GAGCGAGGGC ATAAATACAG CTAGCCCCAG
4579

25 GGGTGGAAACA ACTGTGGGAG TCTTGGGTAC TCGCACCTCT TGGCTTGTT GATGCTCCGC
4639

CAGGAAGGCC ACTTGTGTGT GCGTGTCACT TACTTTTTA GTAACAATTC AGATCCAGTG
4699

30 TAAACTTCCG TTCATTGCTC TCCAGTCACA TGCCCCACT TCCCCACAGG TGAAAGTTTT
4759

TCTGAAGTGT TGGGATTGGT TAAGGTCTTT ATTTGTATTA CGTATCTCCC CAAGTCCTCT
35 4819

GTGCCAGCT GCATCTGTCT GAATGGTGCG TGAAGGCTCT CAGACCTTAC ACACCATTAA
4879

40 GTAAGTTATG TTTTACATGC CCCGTTTTG AGACTGATCT CGATGCAGGT GGATCTCCTT
4939

GAGATCCTGA TAGCCTGTTA CAGGAATGAA GTAAAGGTCA GTTTTTTTG TATTGATTTT
4999

5 CACAGCTTG AGGAACATGC ATAAGAAATG TAGCTGAAGT AGAGGGGACG TGAGAGAAGG
5059

GCCAGGCCGG CAGGCCAACCC CTCCTCCAAT GGAAATTCCC GTGTTGCTTC AACTGAGAC
5119

10 AGATGGACT TAACAGGCAA TGGGGTCCAC TTCCCCCTCT TCAGCATCCC CCGTACC
5176

(2) INFORMATION FOR SEQ ID NO:14:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

25 Met Asn Leu Glu Gly Gly Arg Gly Glu Phe Gly Met Ser Ala
1 5 10 15

Val Ser Cys Gly Asn Gly Lys Leu Arg Gln Trp Leu Ile Asp Gln Ile
20 25 30

30 Asp Ser Gly Lys Tyr Pro Gly Leu Val Trp Glu Asn Glu Glu Lys Ser
35 40 45

Ile Phe Arg Ile Pro Trp Lys His Ala Gly Lys Gln Asp Tyr Asn Arg
35 50 55 60

Glu Glu Asp Ala Ala Leu Phe Lys Ala Trp Ala Leu Phe Lys Gly Lys
65 70 75 80

40 Phe Arg Glu Gly Ile Asp Lys Pro Asp Pro Pro Thr Trp Lys Thr Arg
85 90 95

Leu Arg Cys Ala Leu Asn Lys Ser Asn Asp Phe Glu Glu Leu Val Glu
100 105 110

Arg Ser Gln Leu Asp Ile Ser Asp Pro Tyr Lys Val Tyr Arg Ile Val
5 115 120 125

Pro Glu Gly Ala Lys Lys Gly Ala Lys Gln Leu Thr Leu Glu Asp Pro
130 135 140

Gln Met Ser Met Ser His Pro Tyr Thr Met Thr Pro Tyr Pro Ser
10 145 150 155 160

Leu Pro Ala Gln Gln Val His Asn Tyr Met Met Pro Pro Leu Asp Arg
165 170 175

Ser Trp Arg Asp Tyr Val Pro Asp Gln Pro His Pro Glu Ile Pro Tyr
15 180 185 190

Gln Cys Pro Met Thr Phe Gly Pro Arg Gly His His Trp Gln Gly Pro
20 195 200 205

Ala Cys Glu Asn Gly Cys Gln Val Thr Gly Thr Phe Tyr Ala Cys Ala
210 215 220

Pro Pro Glu Ser Gln Ala Pro Gly Val Pro Thr Glu Pro Ser Ile Arg
25 225 230 235 240

Ser Ala Glu Ala Leu Ala Phe Ser Asp Cys Arg Leu His Ile Cys Leu
245 250 255

Tyr Tyr Arg Glu Ile Leu Val Lys Glu Leu Thr Thr Ser Ser Pro Glu
30 260 265 270

Gly Cys Arg Ile Ser His Gly His Thr Tyr Asp Ala Ser Asn Leu Asp
35 275 280 285

Gln Val Leu Phe Pro Tyr Pro Glu Asp Asn Gly His Arg Lys Asn Ile
290 295 300

Glu Asn Leu Leu Ser His Leu Glu Arg Gly Val Val Leu Trp Met Ala
40 305 310 315 320

Pro Asp Gly Leu Tyr Ala Lys Arg Leu Cys Gln Ser Thr Ile Tyr Trp
325 330 335

Asp Gly Pro Leu Ala Leu Cys Asn Asp Arg Pro Asn Lys Leu Glu Arg
5 340 345 350

Asp Gln Thr Cys Lys Leu Phe Asp Thr Gln Gln Phe Leu Ser Glu Leu
355 360 365

10 Gln Ala Phe Ala His His Gly Arg Ser Leu Pro Arg Phe Gln Val Thr
370 375 380

Leu Cys Phe Gly Glu Glu Phe Pro Asp Pro Gln Arg Gln Arg Lys Leu
385 390 395 400

15 Ile Thr Ala His Val Glu Pro Leu Leu Ala Arg Gln Leu Tyr Tyr Phe
405 410 415

20 Ala Gln Gln Asn Ser Gly His Phe Leu Arg Gly Tyr Asp Leu Pro Glu
420 425 430

His Ile Ser Asn Pro Glu Asp Tyr His Arg Ser Ile Arg His Ser Ser
435 440 445

25 Ile Gln Glu
450

(2) INFORMATION FOR SEQ ID NO:15:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 152 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: DNA (genomic)

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TTTTCTCTAC AGTCACCTCC CTGTTTACCA AAGATAATCA CAATAAGTCC AGTTTACTTA
60

5 CAAAACAAGT TTAGTTATTA GAGGAAACTA AAACCTCAGG ATTCAAGTCCA GATAATT
120

AAAAACTCTA AAACAATGGA CAGGGCTAGA AT
152

10 (2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 152 base pairs
- (B) TYPE: nucleic acid
- 15 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: other nucleic acid

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

25 TGGGCTCGGC CTGGTGGGGC AGCCACAGCG GGACGCAGTA GTGAAAGTCC AGTTTACTTA
60

30 CAAAACAAGT TTAGTTATTA GAGGAAACTA AAACCTCAGG ATTCAAGCAGG GCATGAGGAG
120

30 GCAGCTCCTC ACCCTCCCTT TCTCTTTGT AC
152

35 (2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 152 base pairs
- (B) TYPE: nucleic acid
- 40 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

5

TGGGCTCGGC CTTGGTGGGG CAGCCACAGC GGGACGCAAG TAGTGAGGGC ACTCAGAACG
60

10

CCACTCAGCC CCGACAGGGC ACTCAGAACG CCACTCAGCC CCGACAGGCA GGGCACGAGG
120

AGGCAGCTCC TCACCCCTCCC TTTCTCTTT GT

152

15